

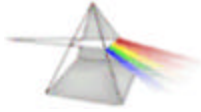
Making Metadata More Accessible, Making Metadata More Expressive

Carl Lagoze

lagoze@cs.cornell.edu

DLI2 PIs Meeting, Roanoke, 2001

- The world according to **Arms**:
 - *Metadata is expensive, we don't need it*
 - *We've spent too much on metadata (and metadata meetings)*
- The world according to **Lesk**:
 - *I dislike funding metadata projects, because they always recommend that every other project should be more expensive*



Projects



- Project Prism (DLI2)
 - <http://www.prism.cornell.edu>
- Harmony (DSTC/JISC/NSF)
 - <http://metadata.net/harmony>
 - Participants
 - Dan Brickley (ILRT)
 - Jane Hunter (DSTC)
 - Carl Lagoze (Cornell)
 - Libby Miller (ILRT)

Some Goals/Hypotheses

- Is common simple metadata useful? How does it relate to more complex metadata? How to model metadata for more complex purposes?
- How to people want to use metadata?
 - Research discovery
 - Content management, etc. (e.g., preservation, authentication, provenance)
- What is the nature of the metadata that people are creating?

Understanding and Evaluating Metadata based on Query Capabilities

- Simple boolean queries on tags?
- Agent, time, place questions?
 - *Who was responsible for what and when*
 - *Preservation, provenance, trust, rights management*

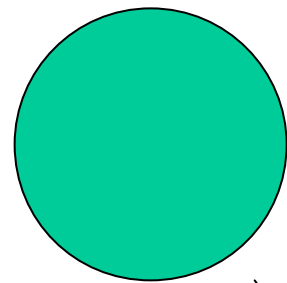
Dublin Core Metadata Initiative

- Simplicity vs. Complexity
 - C. Lagoze, “Keeping Dublin Core Simple: Cross Domain Discovery or Resource Description?,” *D-Lib Magazine*, 7 (1), 2001.
- Can "semantic interoperability" be achieved through such imprecise "standards"
 - Tags provide a useful context for automatic indexing?
 - Evidence is not good (CIMI experiments)

Open Archives Initiative

- <http://www.openarchives.org>
- KISS as a research topic?
- Technical specification facilitates research goals
 - Simple protocol for exposing lots of metadata
 - Mandated metadata format (DC) allows us to study:
 - How people use/abuse "simple" metadata
 - How people dumb-down complex semantics ("graceful degradation")
 - Extensibility facilities allow us to study:
 - How people use metadata for unique community needs
 - How metadata is useful for creating variant services

Resource-centric Metadata



?

Title	Anna Karenina
Author	Leo Tolstoy
Illustrator	Orest Vereisky
Translator	Margaret Wettlin
Date Created	1877
Date Translated	1978
Description	Adultery & Depression

Birthplace	Moscow
Birthdate	1828

ABC/Harmony Event-aware metadata model

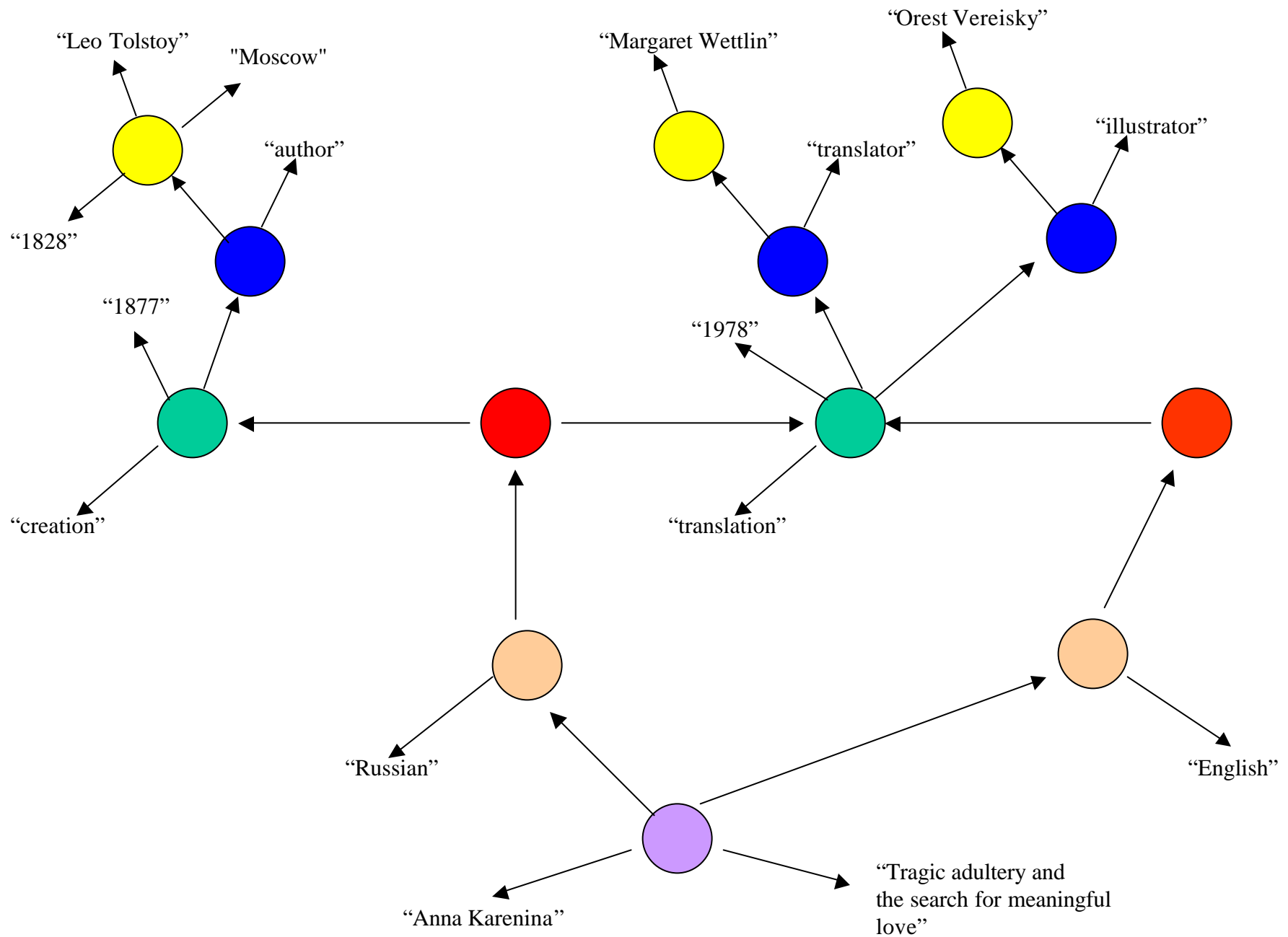
- Goals: better understanding about what we are trying to do with metadata, guidance
- Recognizing inherent lifecycle aspects of description (esp. of digital content)
- Modeling incorporates time (events and situations) as first-class objects
 - Supplies clear attachment points for agents, roles, occurrent properties
- Resource description as a “**story-telling**” activity
- <http://metadata.net/harmony/ABCV2.htm>

Museum Collaboration

- CIMI Collaboration
 - AMOL (Australia Museums and Galleries Online)
 - Natural History Museum of London
 - National History Museum of Denmark
 - Library of Congress/RLG
 - Records collected at: <http://archive.dstc.edu.au/CIMI/>
- CIDOC collaboration
 - DELOS workshop on ontology harmonization

ABC Model

- Entity/relationship and RDF foundation
 - A world of typed and identified entities,
 - that are interrelated in multiple, rich ways
- Core ontology (agents, events, states, actions)
- Time and transformation as first-class concepts
 - Duality of Objects - properties (*adjectives*) that are "constant" and time-dependent (*stateful*) – occurrents vs. continuants
 - Agents and their actions (*verbs*) occur in *events* that modify the state of objects
- <http://metadata.net/harmony/ABCV2.htm>



Queries over RDF Graphs

Rudolf Squish – <http://swordfish.rdfweb.org/rdfquery>

AMOL Demo –

<http://swordfish.rdfweb.org/discovery/2001/03/demos/image.html>

List details of events where Lagoze is a participating agent

```
SELECT ?title, ?type, ?time, ?place, ?name
```

```
FROM
```

```
  http://ilrt.org/discovery/harmony/oai.rdf
```

```
WHERE
```

```
  (web::type ?event abc::Event)
```

```
  (abc::context ?event ?context)
```

```
  ....
```

```
AND ?name ~ lagoze
```

```
USING web FOR http://www.w3.org/1999/02/22-rdf-syntax-ns#
```

Metadata for complex multimedia objects

- Semi-automated generation of SMIL mark-up of distance learning objects (video, slides, transcripts)
- Developed metadata model and web search interface to enable searches across
 - Institutions (Cornell and DSTC);
 - Collections within an institution;
 - Within individual presentations.

<http://sunspot.dstc.edu.au:8888/smil/search.html>

- Paper at ECDL



Search

[BROWSE PRESENTATIONS](#)

Search Archive

[SEARCH TIPS](#)

Enter Search word/s (separating with spaces) and retrieve matching SMIL Presentation slide segments

(e.g., *Project, lipids, plastic information environments*)

Search is case insensitive

Search

--Search DSTC Archive--
--Search Cornell CS501 Archive--
Biosciences Lecture 1
Biosciences Lecture 10
CS182 Lecture

Match on ANY of the given words ☐

Include word variants (e.g. lipid/lipids/phospholipids etc) ☐

☒ Search by Keywords

Limit results to

☐ Search on Full Text (includes keywords)

matches.

Etc.

- Bibliographic community – events and new cataloging models
 - C. Lagoze, “Business Unusual; How "event awareness" may breathe life into the catalog,” presented at Bicentennial Conference on Bibliographic Control in the New Millennium, Library of Congress, Washington DC, 2000.